



Thank you, Gerard. Gerard ordinarily lives nine timezones and 9000 miles away from me, so my midnight is his mid-morning, and that's when he always manages to call me with the latest big news about WZ. But if it weren't for Gerard and his big news and midnight phone calls, I don't think WiktionaryZ would be where it is.

WZ is one of those projects that can't be done, or at least that's what some people said when we began. It isn't done yet, but we have a good start, and I'd like to tell you about it.

Could I see by a show of hands who here has a printed defining dictionary? A translating dictionary? A translating dictionary between two languages that aren't English? The reference section in my library has a book that takes 1000 common English words and offers translations into some 20 languages. The words are in alphabetical order by the English. If you happened to speak Dutch as a native language, you'll have to study English or this book will do you no good.

Translating dictionaries

Fine for common languages, common words.

When you look at the existing Wiktionaries, you'll see that there are several of them. In fact, there is one for each language, or at least one for each language that anybody has bothered to start yet. There's an English Wiktionary, a French Wiktionary, a Dutch Wiktionary, Swedish, Chinese, and so on. Unlike that dusty tome that you use for a doorstep, the English Wiktionary aims to describe not only words in English but all words in all languages. It is the English Wiktionary because it only describes those words in English. The word may be in Finnish, but the definition or translation of it will be in English. We'd like to have all the English words translated into every language. That's already a big job, but we're gradually getting it done.

All words in all languages

- ISO 639-3: 7602 languages
- Average vocabulary: 60,000 words – 100,000 words
- OED: 300,000 words; 600,000 word forms
- Wikis can be that big.

In fact, wikis may be the only hope for accomplishing a task this huge.

Wikis have

*The capacity to store this much data

*The people to help maintain it. Lexicographic data can be costly to maintain because of its size.

*The means to make it accessible.

The problem

- A lot of redundant effort for one meaning

English definition in English Wiktionary	Spanish translation in English Wiktionary
English translation in Spanish Wiktionary	Spanish definition in Spanish Wiktionary

“paper” – a sheet of would be a good example to discuss
You now have four articles for just two words.

The problem

- A lot of redundant effort for one meaning
- It just gets worse!

English definition in English Wiktionary	Spanish translation in English Wiktionary	French translation in English Wiktionary
English translation in Spanish Wiktionary	Spanish definition in Spanish Wiktionary	French translation in Spanish Wiktionary
English translation in French Wiktionary	Spanish translation in French Wiktionary	French definition in French Wiktionary

You now have nine articles just for three words.

That's a lot of places to

Put the data

Maintain the data

Look for the data

Make a mistake

The only things that are not redundant here are the English definition in English, the Spanish definition in Spanish, and the French definition in French. The rest are just relationships being handled really badly.

The problem

- Different senses have different translations.
 - Board: a piece of wood
 - Swedish: bräda
 - Board: a committee
 - Swedish: styrelse
- Words and meaning do not have a one-to-one relationship.
- Languages divide concepts differently.



Formatting doesn't cut it.

Relationships not maintained

Manual process

Flat data structure

Dictionary data is highly structured

Same information in each article

Demo of stupid way that English Wiktionary handles this: big tables, no good clues in translation articles which definition is which.

Grandchild problem and words that don't translate?

Hinge on Defined Meaning

- Translations create articles
- Translations remain associated with a meaning.
- Definition shows in the user's language

WiktionaryZ:board

Your user interface language preference: **en** - [set your preferences](#)

– Language: English

– **A piece of wood or similar material that has been sawn into a regular shape, usually in preparation ...**

– Definition

Language	Text
<i>English</i>	A piece of wood or similar material that has been sawn into a regular shape, usually in preparation for using it in some sort of construction.
<i>Dutch</i>	Een stuk hout of gelijkwaardig materiaal dat in een regelmatige vorm is gezaagd, meestal ter voorbereiding om het te gebruiken voor een bouwwerk.

+ Alternative definitions

– Synonyms and translations

Expression	Language	Spelling	Identical meaning?
	<i>Dutch</i>	plank	<input checked="" type="checkbox"/>

Defined Meaning

- A definition.
- Translate that definition.
- Add translations.
- Add synonyms.
- Translations link automatically.

Defined Meaning	Synonyms
Very large	Huge, enormous, gigantic
Très grand	Énorme, immense, gigantesque
Muy grande	Enorme, gigante

Semantic drift is the inconsistency that happens when

-somebody changes one of these definitions after the fact

Or

-somebody translates a definition incorrectly or too loosely

Links can be created that nobody knew.

*****Demo goes *here*.*****

mayor: comparativo de grande; más grande

English translation of definition: comparative of grande (large); more large

English translation of word: bigger; larger

English, mayor: the leader of a city or municipality

Spanish: el líder de una ciudad o un municipio

Spanish translation of word: alcalde

So what?

- See any language combination
- Ready for re-use:
 - Machine translation
 - Spelling and grammar checkers
 - Translation glossaries
 - Custom language-learning drills
- Instant Commons
- Things we haven't invented

Somebody translates the French to Thai. Somebody else translates the English to Welsh. Somebody else links the French to the English.

Need a Welsh-Thai translation? Click the button.

Combinations no publisher would touch.

Information maintains its relationships and the computer knows what each piece is and what it does.

Gerard's definition of success for the project is when somebody does something with it that we haven't thought of yet! What that means is that we have set it up correctly enough and created or imported enough content that somebody can do something novel and unanticipated with it.

Multilingual ontology: stores relationships between words, across languages. Useful for indexing of other data such as images.

Still under construction

- Language-specific attributes: part of speech, gender
- Wikilinks
- Categories
- Way to merge extra Defined Meanings

Need to be very careful how we implement these

What WiktionaryZ is not

- An attempt to close down or destroy existing Wiktionaries
- An English- or European-centric effort
- An effort only by volunteers: partners and specialized content

If communities fold their projects into WZ, it's because WZ merits it and the community chooses to do so.

Efforts are continuing to incorporate special character sets and coexist with languages that read right-to-left.

Each language will have a community for language-specific issues with an International Beer Parlour for overall discussions in any language.

And it's not only Gerard's pet project, either!

We hope to bring in people from all languages, especially those for whom editing entirely in a second or third language would be challenging.

We do want data to go both ways, between *.wikt and WZ. There will be bots.

Licensing presently makes this transfer one-way from WZ to *.wikt

Try this at home

- Create a user account.
<http://wiktionaryz.org/>
- Add your Babel templates.
- Understand Defined Meaning.
- Ask a bureaucrat for edit rights.
- Try it and tell us what you think!
- Be patient.

Babel	
fra	Cet utilisateur a pour langue maternelle le français.
eng-3	This user is able to contribute with an advanced level of English.
spa-2	Este usuario puede contribuir con un nivel intermedio de español.
deu-1	Dieser Benutzer hat Grundkenntnisse der deutschen Sprache.
cmn-1	该用户能以基本的官话进行交流。 該用戶能以基本的官話進行交流。
lat-1	Hic ususarius simplice latinitate contribuere potest.

Search user languages

It's not open to the public yet because it's still under construction.

Please report bugs and discuss concerns/problems.

If you're interested in helping with programming, let us know!